Improved Estimation for Temporally Clustered Extremes

Lee Fawcett and Dave Walshaw

Newcastle University, Newcastle-upon-Tyne, U.K.

lee.fawcett@ncl.ac.uk

18th TIES Conference: Mikulov, Czech Republic, August 2007

Structure of this talk

1. **Background and motivation**
   - Why extreme value theory?
   - Threshold methods

Structure of this talk

1. **Background and motivation**
   - Why extreme value theory?
   - Threshold methods
2. **Example: Oceanographic data**
   - The data
   - Temporal dependence
   - Return levels

Structure of this talk

1. **Background and motivation**
    - Why extreme value theory?
    - Threshold methods
2. **Example: Oceanographic data**
    - The data
    - Temporal dependence
    - Return levels
3. **Simulation study**

Why extreme value theory?

- As an abstract study of random phenomena, the subject can be traced back to the **early 20th century**

Why extreme value theory?

- As an abstract study of random phenomena, the subject can be traced back to the **early 20th century**
- It was not until the **1950s** that the methodology was proposed for modelling genuine physical phenomena

Why extreme value theory?

- As an abstract study of random phenomena, the subject can be traced back to the **early 20th century**
- It was not until the **1950s** that the methodology was proposed for modelling genuine physical phenomena
- *Early applications*: Civil engineering – e.g. structural design – extreme value theory provided a framework in which an estimate of anticipated forces could be made using historical data

Why extreme value theory?




Why extreme value theory?

- Extremes are scarce – thus, estimates are often required for levels of a process that are much greater than have already been observed

Why extreme value theory?

- Extremes are scarce – thus, estimates are often required for levels of a process that are much greater than have already been observed
- This implies an extrapolation from observed to unobserved levels ...

Why extreme value theory?

- Extremes are scarce – thus, estimates are often required for levels of a process that are much greater than have already been observed
- This implies an extrapolation from observed to unobserved levels ...
- ... extreme value theory provides a class of models to enable such extrapolation

Why extreme value theory?

- Extremes are scarce – thus, estimates are often required for levels of a process that are much greater than have already been observed
- This implies an extrapolation from observed to unobserved levels ...
- ... extreme value theory provides a class of models to enable such extrapolation
- With no empirical or physical basis, asymptotic argument is used to develop extreme value models ...

Why extreme value theory?

- Extremes are scarce – thus, estimates are often required for levels of a process that are much greater than have already been observed
- This implies an extrapolation from observed to unobserved levels ...
- ... extreme value theory provides a class of models to enable such extrapolation
- With no empirical or physical basis, asymptotic argument is used to develop extreme value models ...
- ... however, we are better–off using techniques that at least have *some* sort of rationale!

Threshold methods

“Block maxima” approach

Threshold methods

“Block maxima” approach

- e.g. *annual maximum wind speeds* might be modelled by an appropriate limiting distribution, such as the **generalised extreme value distribution**

Threshold methods

"Block maxima" approach

- e.g. *annual maximum wind speeds* might be modelled by an appropriate limiting distribution, such as the **generalised extreme value distribution**
- Highly inefficient!

Threshold methods

"Block maxima" approach

- e.g. *annual maximum wind speeds* might be modelled by an appropriate limiting distribution, such as the **generalised extreme value distribution**
- Highly inefficient!

Threshold methods

Threshold methods

"Block maxima" approach

- e.g. *annual maximum wind speeds* might be modelled by an appropriate limiting distribution, such as the **generalised extreme value distribution**
- Highly inefficient!

Threshold methods

- An observation is extreme if it exceeds some high cut–off point (**threshold**)

Threshold methods

"Block maxima" approach

- e.g. *annual maximum wind speeds* might be modelled by an appropriate limiting distribution, such as the **generalised extreme value distribution**
- Highly inefficient!

Threshold methods

- An observation is extreme if it exceeds some high cut–off point (**threshold**)
- Use *all* observations which exceed this cut–off point – i.e. use *all* extremes!

Threshold methods

The generalised Pareto distribution (GPD)

Under very broad conditions, if it exists, any limiting distribution as $u \to \infty$ of $(X-u)|X>u$ is of GPD form, where

$$G(y;\sigma,\xi) \;=\; 1-\left(1+\frac{\xi y}{\sigma}\right)_+^{-1/\xi},$$

where $a_+ = \max(0,a)$ and σ $(\sigma > 0)$ and ξ $(-\infty < \xi < \infty)$ are **scale** and **shape** parameters (respectively).

Threshold methods

Practical implementation

Threshold methods

Practical implementation

1. Choose some threshold u which is high enough so that the GPD is a good model for $(X - u | X > u)$

Threshold methods

Practical implementation

1. Choose some threshold u which is high enough so that the GPD is a good model for $(X - u|X > u)$
2. Fit the GPD to the observed excesses $x - u$

Threshold methods

Practical implementation

1. Choose some threshold u which is high enough so that the GPD is a good model for $(X - u|X > u)$
2. Fit the GPD to the observed excesses $x - u$
3. Use the fitted GPD to provide estimates of extreme quantiles, or **return levels** (*see later*)

Threshold methods

Problems?

Threshold methods

Problems?

- Short–range serial correlation, almost always present in environmental time series, can no longer be ignored

Threshold methods

Problems?

- Short–range serial correlation, almost always present in environmental time series, can no longer be ignored
- What do we do about this?

Threshold methods

Problems?

- Short–range serial correlation, almost always present in environmental time series, can no longer be ignored
- What do we do about this?

'**remove**' it?

Threshold methods

Problems?

- Short–range serial correlation, almost always present in environmental time series, can no longer be ignored
- What do we do about this?

‘**remove**’ it? **ignore** it?

Threshold methods

Problems?

- Short–range serial correlation, almost always present in environmental time series, can no longer be ignored
- What do we do about this?

‘**remove**’ it? **ignore** it? **model** it?

Threshold methods

Problems?

- Short–range serial correlation, almost always present in environmental time series, can no longer be ignored
- What do we do about this?

  ‘**remove**’ it? **ignore** it? **model** it?

- **How** do we choose to ‘remove’ it?

Threshold methods

Problems?

- Short–range serial correlation, almost always present in environmental time series, can no longer be ignored
- What do we do about this?

  '**remove**' it? **ignore** it? **model** it?

- **How** do we choose to 'remove' it? **Can** we simply ignore it?

Threshold methods

Problems?

- Short–range serial correlation, almost always present in environmental time series, can no longer be ignored
- What do we do about this?

  '**remove**' it? **ignore** it? **model** it?

- **How** do we choose to 'remove' it? **Can** we simply ignore it? **How** do we model it?

The data

- A series of 3–hourly measurements on sea–surge were obtained from Newlyn, southwest England, collected over a three year period

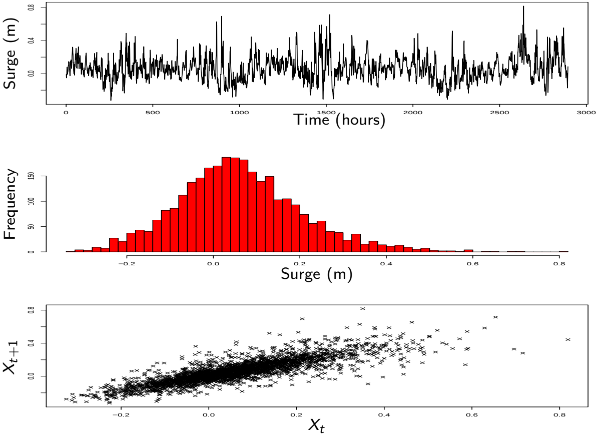
The data

- A series of 3–hourly measurements on sea–surge were obtained from Newlyn, southwest England, collected over a three year period
- Sea–surge is the meteorologically–induced non–tidal component of the still–water level of the sea

The data

- A series of 3–hourly measurements on sea–surge were obtained from Newlyn, southwest England, collected over a three year period
- Sea–surge is the meteorologically–induced non–tidal component of the still–water level of the sea
- Practical motivation: **structural failure** – probably a sea–wall in this case – is likely under the condition of extreme surges

The data



The data

The data

The main problem with implementing the GPD for excesses above a high threshold is the presence of substantial serial correlation in the sequence of 3–hourly surges.

The data

The main problem with implementing the GPD for excesses above a high threshold is the presence of substantial serial correlation in the sequence of 3–hourly surges.

So what do we do?

The data

The main problem with implementing the GPD for excesses above a high threshold is the presence of substantial serial correlation in the sequence of 3–hourly surges.

So what do we do?

'**remove**' it?

The data

The main problem with implementing the GPD for excesses above a high threshold is the presence of substantial serial correlation in the sequence of 3–hourly surges.

So what do we do?

'**remove**' it? **ignore** it?

The data

The main problem with implementing the GPD for excesses above a high threshold is the presence of substantial serial correlation in the sequence of 3–hourly surges.

So what do we do?

'**remove**' it? **ignore** it? **model** it?

Runs declustering

1. Choose a value for the **declustering parameter** κ

Runs declustering

1. Choose a value for the **declustering parameter** κ
2. A 'cluster' of threshold excesses is then deemed to have terminated as soon as at least κ consecutive observations fall below the threshold

Runs declustering

1. Choose a value for the **declustering parameter** κ
2. A 'cluster' of threshold excesses is then deemed to have terminated as soon as at least κ consecutive observations fall below the threshold
3. Select a single observation from each cluster to represent that cluster, and model the set of selected cluster inhabitants

Runs declustering

- But how do we choose κ?

Runs declustering

- But how do we choose κ?
- Which observation should be used to 'represent' each cluster?

Runs declustering

- But how do we choose κ?
- Which observation should be used to 'represent' each cluster?
  - "**Peaks Over Thresholds**" (**Davison and Smith, 1990**)

Runs declustering

- But how do we choose κ?
- Which observation should be used to 'represent' each cluster?
  - "**Peaks Over Thresholds**" (**Davison and Smith, 1990**)
  - Some other cluster inhabitant?

Runs declustering

- But how do we choose κ?
- Which observation should be used to 'represent' each cluster?
  - "**Peaks Over Thresholds**" (**Davison and Smith, 1990**)
  - Some other cluster inhabitant?
- Still wasteful of precious extremes!

Runs declustering

- But how do we choose κ?
- Which observation should be used to 'represent' each cluster?
  - "**Peaks Over Thresholds**" (**Davison and Smith, 1990**)
  - Some other cluster inhabitant?
- Still wasteful of precious extremes!
- And why runs declustering anyway?

Ignoring temporal dependence

Declustering seems like such a hassle... what's wrong with ignoring temporal dependence and just using all threshold excesses anyway?

Ignoring temporal dependence

Declustering seems like such a hassle... what's wrong with ignoring temporal dependence and just using all threshold excesses anyway?

- Maximum likelihood estimators will be **unbiased**

Ignoring temporal dependence

Declustering seems like such a hassle... what's wrong with ignoring temporal dependence and just using all threshold excesses anyway?

- Maximum likelihood estimators will be **unbiased**
- However, **standard errors** will be **underestimated**

Ignoring temporal dependence

Declustering seems like such a hassle... what's wrong with ignoring temporal dependence and just using all threshold excesses anyway?

- Maximum likelihood estimators will be **unbiased**
- However, **standard errors** will be **underestimated**

Idea: Initially ignore dependence, but then *adjust the standard errors*?

Adjusting to account for dependence (Smith, 1991)

- Assuming independence, denote the observed information matrix by $\mathbf{H}$

Adjusting to account for dependence (Smith, 1991)

- Assuming independence, denote the observed information matrix by $\mathbf{H}$
- If independence were a valid assumption, then the variance–covariance matrix would be (approximately) $\mathbf{H}^{-1}$

Adjusting to account for dependence (Smith, 1991)

- Assuming independence, denote the observed information matrix by $\mathbf{H}$
- If independence were a valid assumption, then the variance–covariance matrix would be (approximately) $\mathbf{H}^{-1}$
- To account for dependence, replace $\mathbf{H}^{-1}$ with $\mathbf{H}^{-1}\mathbf{V}\mathbf{H}^{-1}$, where $\mathbf{V}$ is the variance–covariance matrix of the likelihood gradient vector

Adjusting to account for dependence (Smith, 1991)

- Assuming independence, denote the observed information matrix by $\mathbf{H}$
- If independence were a valid assumption, then the variance–covariance matrix would be (approximately) $\mathbf{H}^{-1}$
- To account for dependence, replace $\mathbf{H}^{-1}$ with $\mathbf{H}^{-1}\mathbf{V}\mathbf{H}^{-1}$, where $\mathbf{V}$ is the variance–covariance matrix of the likelihood gradient vector
- $\mathbf{V}$ can be estimated by decomposing the log–likelihood sum into its contributions by year, say, which are independent

Adjusting to account for dependence (Smith, 1991)

Similar arguments (not given here) can be applied to modify the procedure for testing hypotheses.

Adjusting to account for dependence (Smith, 1991)

Similar arguments (not given here) can be applied to modify the procedure for testing hypotheses.

Technique:

Adjusting to account for dependence (Smith, 1991)

Similar arguments (not given here) can be applied to modify the procedure for testing hypotheses.

Technique:

- Replace the critical value used to carry out a likelihood ratio test by a larger value

Adjusting to account for dependence (Smith, 1991)

Similar arguments (not given here) can be applied to modify the procedure for testing hypotheses.

Technique:

- Replace the critical value used to carry out a likelihood ratio test by a larger value
- The significance levels are then adjusted to take the dependence into account

Adjusting to account for dependence (Smith, 1991)

Similar arguments (not given here) can be applied to modify the procedure for testing hypotheses.

Technique:

- Replace the critical value used to carry out a likelihood ratio test by a larger value
- The significance levels are then adjusted to take the dependence into account

Overall effect:

Adjusting to account for dependence (Smith, 1991)

Similar arguments (not given here) can be applied to modify the procedure for testing hypotheses.

Technique:

- Replace the critical value used to carry out a likelihood ratio test by a larger value
- The significance levels are then adjusted to take the dependence into account

Overall effect:

- Inflate the standard errors associated with MLEs

Adjusting to account for dependence (Smith, 1991)

Similar arguments (not given here) can be applied to modify the procedure for testing hypotheses.

Technique:

- Replace the critical value used to carry out a likelihood ratio test by a larger value
- The significance levels are then adjusted to take the dependence into account

Overall effect:

- Inflate the standard errors associated with MLEs
- Increase the width of confidence intervals (obtained directly or via profile likelihood)

So which is best?

So let's compare these two methods of dealing with temporal dependence:

So which is best?

So let's compare these two methods of dealing with temporal dependence:

1. '**Removing**' dependence, and

So which is best?

So let's compare these two methods of dealing with temporal dependence:

1. '**Removing**' dependence, and
2. '**Ignoring**' dependence, but then adjusting inferences accordingly

So which is best?

So let's compare these two methods of dealing with temporal dependence:

1. '**Removing**' dependence, and
2. '**Ignoring**' dependence, but then adjusting inferences accordingly

For **modelling** this dependence, see, for example, **Fawcett and Walshaw (2006)**.

Results

For a suitable high threshold u, and – in the case of peaks over thresholds – using a value of κ which allows for wave propagation time (see **Coles and Tawn, 1991**), the following results for the GPD were obtained:

Results

For a suitable high threshold u, and – in the case of peaks over thresholds – using a value of κ which allows for wave propagation time (see **Coles and Tawn, 1991**), the following results for the GPD were obtained:

	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\lambda}_u$
All excesses	0.104	–0.090	0.059
95% CI	(0.082, 0.126)	(–0.217, 0.037)	(0.058, 0.060)
Cluster peaks	0.187	–0.259	0.013
95% CI	(0.109, 0.265)	(–0.545, 0.027)	(0.012, 0.014)

Return levels

A **return level** z_s, with associated **return period** s, can be thought of as the level which we can expect to be exceeded once every s observations.

Return levels

A **return level** z_s, with associated **return period** s, can be thought of as the level which we can expect to be exceeded once every s observations.

Now,

$$\Pr(X > z_s) \quad = \quad \Pr(X > z_s | X > u)\Pr(X > u)$$

Return levels

A **return level** z_s, with associated **return period** s, can be thought of as the level which we can expect to be exceeded once every s observations.

Now,

$$\begin{aligned}\Pr(X > z_s) &= \Pr(X > z_s | X > u)\Pr(X > u) \\ &= \Pr(X > u)\left[1 - G(z_s - u; \sigma, \xi)\right]\end{aligned}$$

Return levels

A **return level** z_s, with associated **return period** s, can be thought of as the level which we can expect to be exceeded once every s observations.

Now,

$$
\begin{aligned}
\Pr(X > z_s) &= \Pr(X > z_s | X > u)\Pr(X > u) \\
&= \Pr(X > u)\left[1 - G(z_s - u; \sigma, \xi)\right] \\
&= \lambda_u \left[1 + \xi\left(\frac{z_s - u}{\sigma}\right)\right]_+^{-1/\xi},
\end{aligned}
$$

Return levels

A **return level** z_s, with associated **return period** s, can be thought of as the level which we can expect to be exceeded once every s observations.

Now,

$$\begin{aligned}\Pr(X > z_s) &= \Pr(X > z_s | X > u)\Pr(X > u) \\ &= \Pr(X > u)\left[1 - G(z_s - u; \sigma, \xi)\right] \\ &= \lambda_u \left[1 + \xi\left(\frac{z_s - u}{\sigma}\right)\right]_+^{-1/\xi},\end{aligned}$$

where $\lambda_u = \Pr(X > u)$.

Return levels

- An appropriate return level can then be estimated by equating the above to s^{-1} and solving for z_s

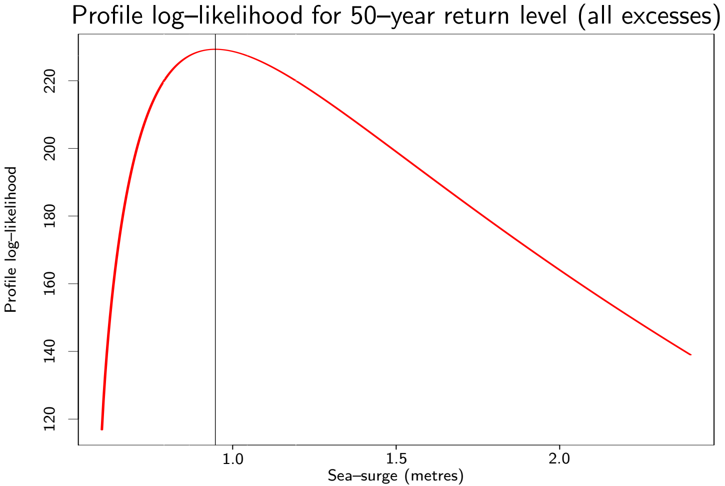
Return levels

- An appropriate return level can then be estimated by equating the above to s^{-1} and solving for z_s
- For confidence intervals for z_s we use **profile likelihood** (**Venzon and Moolgavkar, 1988**) due to the severe asymmetry often encountered in the likelihood surface ...

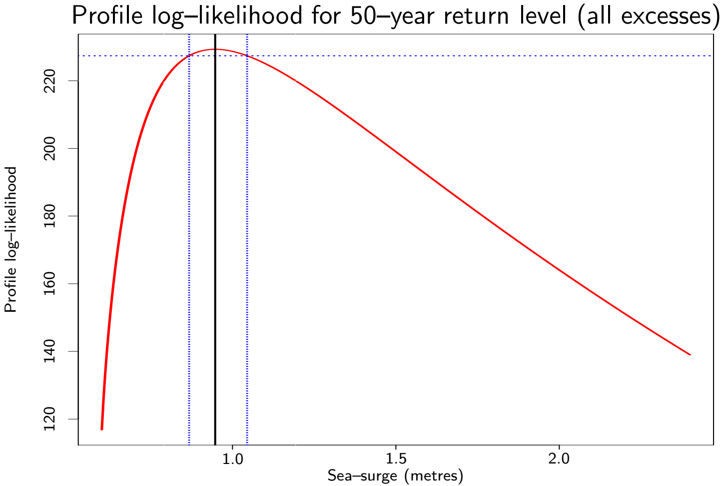
Return levels

- An appropriate return level can then be estimated by equating the above to s^{-1} and solving for z_s
- For confidence intervals for z_s we use **profile likelihood** (**Venzon and Moolgavkar, 1988**) due to the severe asymmetry often encountered in the likelihood surface ...
- ... a modified version of Smith's adjustment is then used on the profile likelihood surface to account for serial dependence

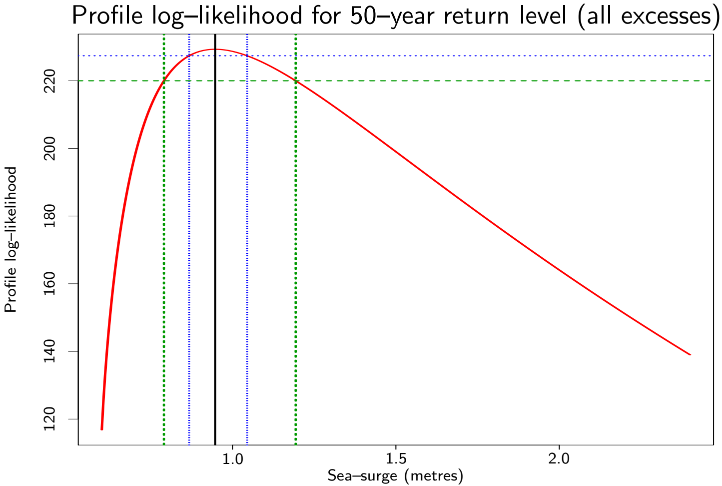
Return levels



Return levels



Return levels



Return levels

The table below shows return level estimates for the **50**, **200** and **1000**–year return periods.

Return levels

The table below shows return level estimates for the **50**, **200** and **1000**–year return periods.

	$\hat{z}_{50}$	$\hat{z}_{200}$	$\hat{z}_{1000}$
All excesses	0.947	1.007	1.068
95% CI	(0.790, 1.193)	(0.844, 1.257)	(0.891, 1.335)
Cluster peaks	0.920	0.951	0.975
95% CI	(0.813, 1.099)	(0.838, 1.008)	(0.858, 1.063)

Return levels

The table below shows return level estimates for the **50**, **200** and **1000**–year return periods.

	$\hat{z}_{50}$	$\hat{z}_{200}$	$\hat{z}_{1000}$
All excesses	0.947	1.007	1.068
95% CI	(0.790, 1.193)	(0.844, 1.257)	(0.891, 1.335)
Cluster peaks	0.920	0.951	0.975
95% CI	(0.813, 1.099)	(0.838, 1.008)	(0.858, 1.063)

There are clear discrepancies between the two approaches... but which estimates should we trust?

Simulated data with extremal dependence

We assume that our series of observations forms a stationary first–order Markov chain.

Simulated data with extremal dependence

We assume that our series of observations forms a stationary first–order Markov chain.

Given a model $f(x_i, x_{i+1}; \underline{\psi})$, $i = 1, \ldots, n-1$, it follows that the joint density function for $x_1, \ldots, x_n$ is given by

$$f(x_1, \ldots, x_n) \quad = \quad \prod_{i=1}^{n-1} f(x_i, x_{i+1}; \underline{\psi}) \Big/ \prod_{i=2}^{n-1} f(x_i; \underline{\psi}).$$

Bivariate threshold excess model

We generate (X_i, X_{i+1}) from a subclass of the complete class of **limiting bivariate extreme value distributions** (various subclasses were tried).

Bivariate threshold excess model

We generate (X_i, X_{i+1}) from a subclass of the complete class of **limiting bivariate extreme value distributions** (various subclasses were tried).

Here, we present the results based on the **logistic** model (**Tawn, 1988**) with **dependence parameter** α.

Bivariate threshold excess model

We generate (X_i, X_{i+1}) from a subclass of the complete class of **limiting bivariate extreme value distributions** (various subclasses were tried).

Here, we present the results based on the **logistic** model (**Tawn, 1988**) with **dependence parameter** α.

- When $\alpha = 1$ we have independence

Bivariate threshold excess model

We generate (X_i, X_{i+1}) from a subclass of the complete class of **limiting bivariate extreme value distributions** (various subclasses were tried).

Here, we present the results based on the **logistic** model (**Tawn, 1988**) with **dependence parameter** α.

- When $\alpha = 1$ we have independence
- Perfect dependence arises from the limit as $\alpha \to 0$

Bivariate threshold excess model

We generate (X_i, X_{i+1}) from a subclass of the complete class of **limiting bivariate extreme value distributions** (various subclasses were tried).

Here, we present the results based on the **logistic** model (**Tawn, 1988**) with **dependence parameter** α.

- When $\alpha = 1$ we have independence
- Perfect dependence arises from the limit as $\alpha \to 0$

For simulation details, see **Fawcett (2005)**.

Study design

1. Set $\underline{\alpha} = (0.01, 0.02, \ldots, 1.00)$

Study design

1. Set $\underline{\alpha} = (0.01, 0.02, \ldots, 1.00)$
2. For $\alpha^{[1]} = 0.01$, simulate a first–order Markov chain (of length *n*) and GPD marginal distribution with known σ and ξ

Study design

1. Set $\underline{\alpha} = (0.01, 0.02, \ldots, 1.00)$
2. For $\alpha^{[1]} = 0.01$, simulate a first–order Markov chain (of length *n*) and GPD marginal distribution with known σ and ξ

Then ...

Study design

3. Use maximum likelihood estimation to fit the GPD to **all threshold excesses**, adjusting for temporal dependence accordingly; use the MLEs for σ and ξ to obtain the MLE for various return levels

Study design

3. Use maximum likelihood estimation to fit the GPD to **all threshold excesses**, adjusting for temporal dependence accordingly; use the MLEs for σ and ξ to obtain the MLE for various return levels
4. Implement runs declustering and fit the GPD to the set of **cluster peak excesses**; again, use the MLEs for σ and ξ to obtain the MLE for various return levels

Study design

3. Use maximum likelihood estimation to fit the GPD to **all threshold excesses**, adjusting for temporal dependence accordingly; use the MLEs for σ and ξ to obtain the MLE for various return levels
4. Implement runs declustering and fit the GPD to the set of **cluster peak excesses**; again, use the MLEs for σ and ξ to obtain the MLE for various return levels
5. Repeat steps (2 – 4) N times to obtain the sampling distributions for σ, ξ and z_s, and compare these to the known values for these parameters

Study design

3. Use maximum likelihood estimation to fit the GPD to **all threshold excesses**, adjusting for temporal dependence accordingly; use the MLEs for σ and ξ to obtain the MLE for various return levels
4. Implement runs declustering and fit the GPD to the set of **cluster peak excesses**; again, use the MLEs for σ and ξ to obtain the MLE for various return levels
5. Repeat steps (2 – 4) N times to obtain the sampling distributions for σ, ξ and z_s, and compare these to the known values for these parameters
6. Repeat steps (2 – 5) for $\alpha^{[2]} = 0.02, \alpha^{[3]} = 0.03, \ldots$

Study design

- Five different values of ξ were used: –0.4, –0.1, 0, 0.3 and 0.8

Study design

- Five different values of ξ were used: –0.4, –0.1, 0, 0.3 and 0.8
- The scale parameter σ was held unit constant. However:

Study design

- Five different values of ξ were used: –0.4, –0.1, 0, 0.3 and 0.8
- The scale parameter σ was held unit constant. However:
  - By construction, *all* observations follow a GPD(σ, ξ)

Study design

- Five different values of ξ were used: –0.4, –0.1, 0, 0.3 and 0.8
- The scale parameter σ was held unit constant. However:
  - By construction, *all* observations follow a GPD(σ, ξ)
  - Excesses above a threshold u will follow a GPD(σ^*, ξ), where

$$\sigma^* = \sigma + \xi u$$

(**threshold stability property** of the GPD)

Study design

- Five different values of ξ were used: –0.4, –0.1, 0, 0.3 and 0.8
- The scale parameter σ was held unit constant. However:
  - By construction, *all* observations follow a GPD(σ, ξ)
  - Excesses above a threshold u will follow a GPD(σ^*, ξ), where

$$\sigma^* = \sigma + \xi u$$

  (**threshold stability property** of the GPD)

- The threshold u was set at a level such that $G(u) = 0.95$

Study design

- Five different values of ξ were used: –0.4, –0.1, 0, 0.3 and 0.8
- The scale parameter σ was held unit constant. However:
  - By construction, *all* observations follow a GPD(σ, ξ)
  - Excesses above a threshold u will follow a GPD(σ^*, ξ), where

$$\sigma^* = \sigma + \xi u$$

  (**threshold stability property** of the GPD)

- The threshold u was set at a level such that $G(u) = 0.95$
- For declustering, κ was set at 20 observations, in line with the Newlyn analysis

Results: $\alpha = 0.2$

The following table shows results for $\alpha = 0.2$, which reflects the type of strong (extremal) dependence often encountered in environmental time series.

Results: $\alpha = 0.2$

The following table shows results for $\alpha = 0.2$, which reflects the type of strong (extremal) dependence often encountered in environmental time series.

True value:	$\sigma^* = 0.302$	$\xi = -0.4$	$z_{50} = 2.479$	$z_{200} = 2.488$
Estimate using all excesses	0.301 (0.251, 0.351)	–0.413 (–0.507, –0.323)	2.459 (2.378, 2.544)	2.568 (2.382, 2.561)

Results: $\alpha = 0.2$

The following table shows results for $\alpha = 0.2$, which reflects the type of strong (extremal) dependence often encountered in environmental time series.

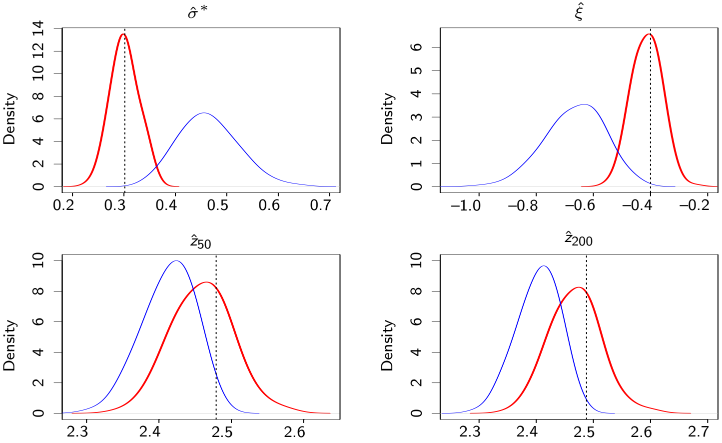
True value:	$\sigma^* = 0.302$	$\xi = -0.4$	$z_{50} = 2.479$	$z_{200} = 2.488$
Estimate using all excesses	0.301 (0.251, 0.351)	–0.413 (–0.507, –0.323)	2.459 (2.378, 2.544)	2.568 (2.382, 2.561)
Estimate using cluster peaks	0.464 (0.358, 0.583)	–0.665 (–0.868, –0.480)	2.404 (2.331, 2.466)	2.413 (2.344, 2.472)

Results: $\alpha = 0.2$

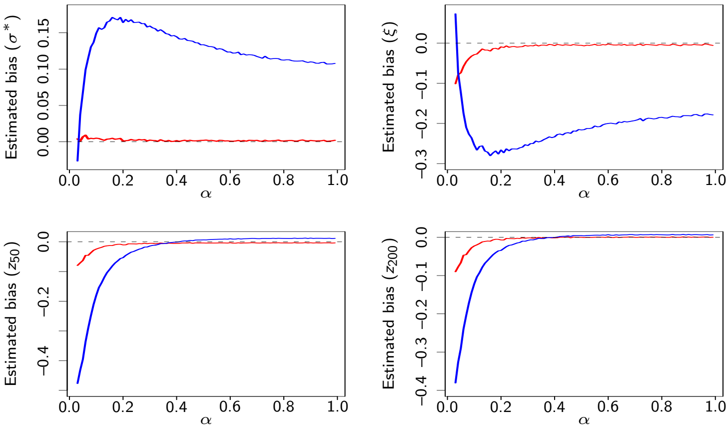
The following table shows results for $\alpha = 0.2$, which reflects the type of strong (extremal) dependence often encountered in environmental time series.

True value:	$\sigma^* = 0.302$	$\xi = -0.4$	$z_{50} = 2.479$	$z_{200} = 2.488$
Estimate using all excesses	0.301 (0.251, 0.351)	–0.413 (–0.507, –0.323)	2.459 (2.378, 2.544)	2.568 (2.382, 2.561)
Estimate using cluster peaks	0.464 (0.358, 0.583)	–0.665 (–0.868, –0.480)	2.404 (2.331, 2.466)	2.413 (2.344, 2.472)
Estimated bias	0.000 (0.162)	–0.013 (–0.265)	–0.020 (–0.066)	–0.020 (–0.084)
MSE	0.001 (0.030)	0.003 (0.080)	0.002 (0.006)	0.002 (0.008)

Results: $\alpha = 0.2$



Results: other levels of dependence



Assessing the adjustment for dependence

		$\hat{\sigma}^*$	$\hat{\xi}$	$\hat{z}_{50}$	$\hat{z}_{200}$
$\alpha = \mathbf{0.2}$	St. dev.	0.039	0.075	0.041	0.047
	mean(e.s.e.)	0.024	0.053	0.019	0.026
	mean(*adjusted* e.s.e.)	0.038	0.076	0.040	0.045
$\alpha = \mathbf{0.5}$	St. dev.	0.026	0.056	0.025	0.030
	mean(e.s.e.)	0.023	0.051	0.019	0.026
	mean(*adjusted* e.s.e.)	0.026	0.056	0.024	0.030
$\alpha = \mathbf{0.8}$	St. dev.	0.024	0.051	0.021	0.025
	mean(e.s.e.)	0.023	0.050	0.019	0.024
	mean(*adjusted* e.s.e.)	0.024	0.051	0.021	0.025

Robustness of results

Declustering scheme	$\sigma^* = 0.302$	$\xi = -0.4$	$z_{50} = 2.479$	$z_{200} = 2.488$
Runs, $\kappa = 30$ hours	0.476	–0.668	2.389	2.397
	(0.423, 0.511)	(–0.766, –0.593)	(2.353, 2.438)	(2.342, 2.428)
Runs, $\kappa = 90$ hours	0.478	–0.680	2.394	2.456
	(0.344, 0.608)	(–0.981, –0.353)	(2.314, 2.482)	(2.397, 2.516)

Robustness of results

Declustering scheme	$\sigma^* = 0.302$	$\xi = -0.4$	$z_{50} = 2.479$	$z_{200} = 2.488$
Runs, $\kappa = 30$ hours	0.476	–0.668	2.389	2.397
	(0.423, 0.511)	(–0.766, –0.593)	(2.353, 2.438)	(2.342, 2.428)
Runs, $\kappa = 90$ hours	0.478	–0.680	2.394	2.456
	(0.344, 0.608)	(–0.981, –0.353)	(2.314, 2.482)	(2.397, 2.516)
Runs, $\kappa = 60$ hours	0.466	–0.621	2.408	2.414
(random excesses)	(0.361, 0.587)	(–0.863, –0.476)	(2.336, 2.471)	(2.345, 2.475)

Robustness of results

Declustering scheme	$\sigma^* = 0.302$	$\xi = -0.4$	$z_{50} = 2.479$	$z_{200} = 2.488$
Runs, $\kappa = 30$ hours	0.476	–0.668	2.389	2.397
	(0.423, 0.511)	(–0.766, –0.593)	(2.353, 2.438)	(2.342, 2.428)
Runs, $\kappa = 90$ hours	0.478	–0.680	2.394	2.456
	(0.344, 0.608)	(–0.981, –0.353)	(2.314, 2.482)	(2.397, 2.516)
Runs, $\kappa = 60$ hours	0.466	–0.621	2.408	2.414
(random excesses)	(0.361, 0.587)	(–0.863, –0.476)	(2.336, 2.471)	(2.345, 2.475)
Blocks	0.409	–0.610	2.365	2.391
(block length = 60 hours)	(0.312, 0.506)	(–0.833, –0.387)	(2.328, 2.402)	(2.354, 2.438)

Robustness of results

Declustering scheme	$\sigma^* = 0.302$	$\xi = -0.4$	$z_{50} = 2.479$	$z_{200} = 2.488$
Runs, $\kappa = 30$ hours	0.476	–0.668	2.389	2.397
	(0.423, 0.511)	(–0.766, –0.593)	(2.353, 2.438)	(2.342, 2.428)
Runs, $\kappa = 90$ hours	0.478	–0.680	2.394	2.456
	(0.344, 0.608)	(–0.981, –0.353)	(2.314, 2.482)	(2.397, 2.516)
Runs, $\kappa = 60$ hours	0.466	–0.621	2.408	2.414
(random excesses)	(0.361, 0.587)	(–0.863, –0.476)	(2.336, 2.471)	(2.345, 2.475)
Blocks	0.409	–0.610	2.365	2.391
(block length $= 60$ hours)	(0.312, 0.506)	(–0.833, –0.387)	(2.328, 2.402)	(2.354, 2.438)
Automatic	0.406	–0.602	2.364	2.389
(Ferro and Segers, 2003)	(0.308, 0.500)	(–0.824, –0.378)	(2.326, 2.404)	(2.350, 2.440)

References

Coles, S.G. and Tawn, J.A. (1991). Modelling extreme multivariate events. *J. R. Statist. Soc., B*, **53**, 377–392.

Davison, A.C. and Smith, R.L. (1990). Models for Exceedances over High Thresholds (with discussion). *J. R. Statist. Soc., B*, **52**, 393–442.

Fawcett, L. (2005). Statistical Methodology for the Estimation of Environmental Extremes. *PhD Thesis*. Newcastle University, Newcastle–upon–Tyne.

Fawcett, L. and Walshaw, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics*, **18**, 173–188.

Fawcett, L. and Walshaw, D. (2006). Markov chain models for extreme wind speeds. *Environmetrics*, **17**, 795–809.

Ferro, C.A.T. and Segers, J. (2003). Inference for clusters of extreme values. *J. R. Statist. Soc., B*, **65**, 545–556.

Smith, R.L. (1991). Regional estimation from spatially dependent data. Preprint. (http:www.stat.unc.edu/postscript/rs/regest.pdf)

Tawn, J.A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika*, **75**, 397–415.

Venzon, D.J. and Moolgavkar, S.H. (1988). Profile-likelihood–based confidence intervals. *Applications of Statistics*, **37**, 87–94.